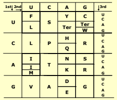


Ancient Transitional Codes Shaped the Genetic Code

Brian K. Davis :: Research Foundation of Southern California, Inc., La Jolla, California ::

Brian K. Davis, D.Sc.
Research Foundation of
Southern California, Inc.
P.O. Box 13595
La Jolla, CA 92039
davis@resfdnsca.org

Introduction

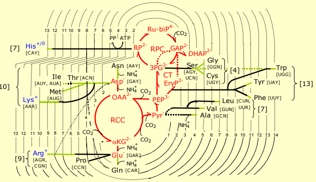
The genetic code of the last common ancestor (LCA), or a minor variant of it, is present in all species. Its origin, in the pre-LCA era, has remained an enigma for four decades. My analysis reveals that all regularities identified in code structure correlate strongly with path-distances in amino acid synthesis. The code accordingly evolved by adding amino acids as they appeared, during the growth of synthesis pathways outward from central metabolism. Codon assignments in the 'universal code' were found to derive from ancient transitional codes, formed deep within the pre-LCA era.

Design of Study

- Amino acid path-distances were evaluated as the number of reaction steps, from citrate cycle, required for synthesis.
- Distinct transition codes were identified and used to reconstruct code evolution.
- The path-distance model was then validated by showing that it unifies over twenty diverse structural regularities identified in the genetic code and pre-LCA proteins.

Amino Acid Synthesis Pathways

- Amino acids form on ancient pathways catalyzed by fifty pre-LCA enzymes.
- The twenty coded amino acids subdivide into four families with precursors OAA, αKG, Pyr, and PEP.

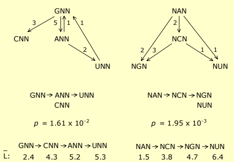


10 Path-length contour
[7] No. pre-LCA enzymes in paths
[CGN] Codons. N, any base, R, purine Y pyrimidine
RCC Reductive citrate cycle
RPC Reductive pentose cycle

Path with multianionic components
Scaled amino acid synthesis path
Unscaled segment of path
Segment with alpha-amino acids
CT Central trunk

Precursor-Product Amino Acid Pairs have Time-Ordered Codons

	Precursor → product	Codons	Path length (L)
1	Asp → Asn	GA• → AA•	1 → 2
2	Asp → Thr	GA• → AC•	1 → 6
3	Asp → Met	GA• → AU•	1 → 7
4	Asp → Lys	GA• → AA•	1 → 10
5	Glu → Gln	GA• → CA•	1 → 2
6	Glu → Pro	GA• → CC•	1 → 4
7	Glu (Asp)	GA• → AG•	1 → 9
	→ Arg	GA• → CG•	1 → 9
8	Gln → His	CA• → CA•	1 → 13
9	Ser → Cys	AG• → UG•	4 → 5
		UC• → UG•	4 → 5
10	Ser → Gly	AG• → GG•	4 → 5
		UC• → GG•	4 → 5
11	Ser → Trp	AG• → UG•	4 → 14
		UC• → UG•	4 → 14
12	Thr → Ile	AC• → AU•	6 → 7



- Extension of amino acid synthesis pathways and code formation were linked as:
  (i) Codons in precursor/product amino acid pairs exhibit time-ordered 5'- and mid-base.
  (ii) Mid-base and mean synthesis path-length (L) correlate strongly among short-path (2-7 steps) amino acids.

Amino Acid and Codon Distribution on Synthesis Pathways

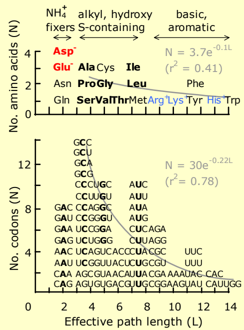
Amino acids with short, medium, and long paths are chemically distinct and encoded differently:

(1) Four NH_4^+ fixers form in 2 steps (or less), include both anionic residues (red), and have codons solely from the NAN column.

(2) Ten amino acids with alkyl, hydroxy and S-bearing side-chains form on 4-7 step paths. Consensus codons for 4-, 5-, and 7-step residues are respectively from the NCN, NGN, and NUN column. 7 of 8 'sets of four' (intact code boxes) encode them.

(3) Six basic (blue) and aromatic residues form on 9-14 step paths. They are encoded mainly by codon doublets.

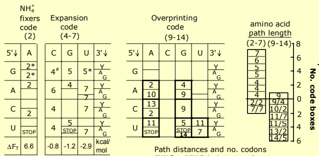
A 14-fold (exponential) fall-off in codon assignments over paths of 4- to 14-steps conforms with gradual slowing in the tempo evolution leading to the 'universal code'.



Bold letters- thermostable amino acids (upper)
- consensus codon mid-base (lower)

Transition Codes Revealed by Amino Acid Path-Distances

- NAN column codons for four '2-step' amino acids (NH_4^+ fixers) and STOP signal are identified as a vestige of the first code. Codons of amino acids with 2-, 4-, 5-, and 7-step paths exhibit column-wise order NAN→NCN→NGN→NUN consistent with column-by-column growth of the code.
- Early amino acids (2-7 step paths) acquired 7 of 8 intact boxes, suggesting each was allotted an intact box of codons on entering the code.
- Latecomer amino acids (9-14 step paths) share 6 of 8 subdivided boxes with an early amino acid (2-7 step path), or stop signal. This indicates they captured codons from early amino acids.



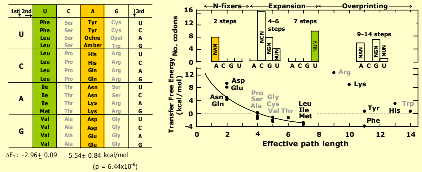
Path-distances and code columns:
p(2-7) = 5.58x10^{-10} (χ^2_{10} = 64.3)
p(9-14) = 0.37 (χ^2_{10} = 10.9)

Path distances and no. codons (NNC = NNU) in intact and subdivided boxes:
p(2-7 vs. 9-14) = 4.63x10^{-5} (χ^2_1 = 16.6)

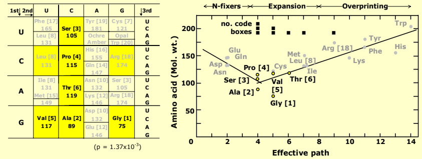
(2), stage of code evolution
*, path-distance advanced 1 step; #, advanced 2 steps
ΔF_T, mean transfer free energy of residues (hydrophobicity) in code column

Path-Distance Model Explains Code Structure

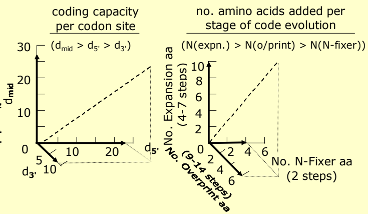
1. Woese (1965): *NAN column triplets code for hydrophilic amino acids, while NUN triplets encode hydrophobic amino acids.*



2. Taylor and Coates (1989): *Six of eight intact boxes code for the smallest amino acids in proteins.*



3. Perlwitz et al. (1988): *Codon mid-base has most coding capacity.*



Conclusion

Amino acid synthesis path length strongly influenced codon acquisition. Analysis of this relation reveals the genetic code evolved in three steps: (i) NH_4^+ Fixers Code, with NAN column triplets, (ii) Code Expansion, and (iii) Overprinting by latecomers. Not only is the genetic code 'universal', therefore, it conserves vestiges of the transition codes that shaped it. tRNA with joint cofactor (amino acid synthesis) and adaptor (translation) functions are implicated in linking code formation with pre-LCA path extension.

Reference

B. K. Davis, 2007. In *Leading-Edge Messenger RNA Research Communications* M. H. Ostrovskiy, Ed. New York: Nova Science, pp 1-32.